Analysis of Volunteer and Automated Water Quality Data: Recommendations for Improvements and Data Management

Campbell Environmental Informatics
Columbia, Maryland
WWW.ENVIRONMENTALINFORMATICS.COM

Jeffrey D. Campbell, Ph.D.
Jeffrey.Campbell@EnvironmentalInformatics.com

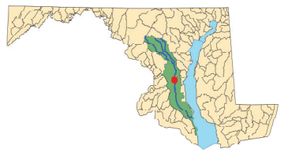
Jug Bay Wetlands Sanctuary
Lothian, Maryland
WWW.JUGBAY.ORG

Overview

- Volunteers observed dissolved oxygen, salinity, pH, temperature, Secchi depth, and collected grab samples for outside lab nutrient analysis.
- Since 2003, data have been recorded at 15 minute intervals from a YSI 6600 located 2 meters away from one of the volunteer sampling sites.
- 128 volunteer observations have been matched with linear interpolation of automated values around volunteer data time.
- DO concentration (mg/l) instead of percent saturation was selected for detailed analysis since the meter measures concentration and performs non-linear computation of percent saturation using additional data.

Jug Bay Wetlands Sanctuary

- Tidal fresh water, Patuxent River
- Middle of 900 sq mi watershed
- Chesapeake Bay National Estuarine Research Reserve, Maryland



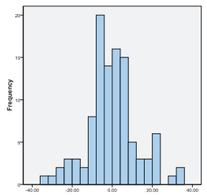
Conclusions & Recommendations

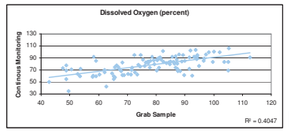
- Data obtained by some volunteers is closer to the automated values than data from other volunteers.
- Dissolved Oxygen values exhibit greatest variation that is consistent with individual differences in collection techniques.
  - Positive bias due to unintended aeration
  - Negative bias due to insufficient water flow at meter probe
- Revise procedures, provide additional training, have periodic refresher training and procedural verification.
- Retain site for data quality evaluation instead of discontinuing this volunteer samples there redundant with automated system.
- Use this site for comparison of the meter to the automated system as part of QA/QC plan for other research projects using the hand-held meter.

Comparison
Volunteer value - automated value

(Difference = Volunteer value - automated value)

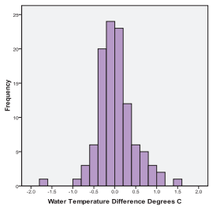
Dissolved Oxygen Saturation needs more analysis

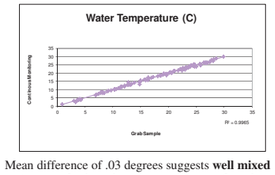




Mean difference (.04) is small but standard deviation of 12.54 gives 95% confidence interval of plus or minus 25. This interval is larger than desired. DO Concentration analyzed in detail to the right.

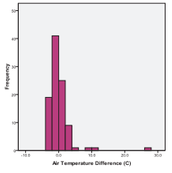
Water Temperature shows excellent consistency

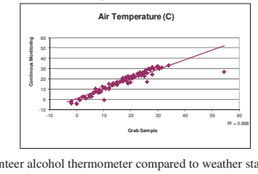




Mean difference of .03 degrees suggests **well mixed water sample** since the automated sensor .3 m above bottom and volunteer samples were from a bucket collected near the surface.

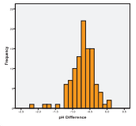
Air Temperature shows unanticipated outliers

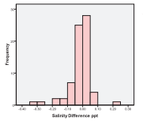




Volunteer alcohol thermometer compared to weather station 475 m away. **Outlier still apparent** even after correcting a volunteer observation measured erroneously in Fahrenheit.

pH and Salinity show methodological limits



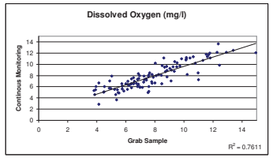


Negative skew of pH differences shows volunteer color interpretation of test strips resulted in a lower pH than the automated meter.

Salinity mean difference of .02 is less than the one decimal place reported on the volunteer's YSI 85 and 95 meter.

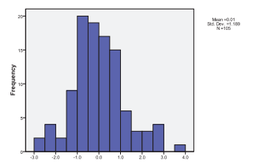
Dissolved Oxygen Concentration Analysis

1) Quick Comparison



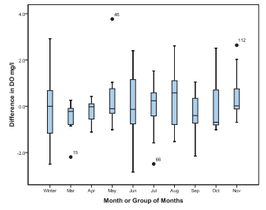
We started with a simple scatter plot comparing volunteer grab samples with the automated samples. Correlation of .76 looks good, but...

2) Differences Histogram



… in this histogram of the differences, we see some differences as large as + 4 and -3 mg/l. Since the average DO value is around 8 mg/l, this seems large. With a standard deviation of 1.19, there is 95% probability that a measurement will be within plus or minus 2.33 mg/l.

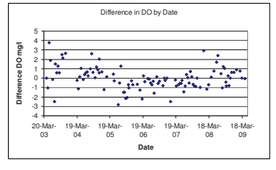
3) Change by Season?



Perhaps there is a seasonal component to the error (cold volunteers rush?) The box plot shows the mean difference, 95% confidence limits and min, max differences and outliers. **No seasonality is obvious.**

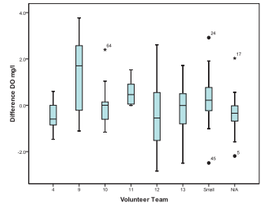
Note: some months are analyzed together because there were too few observations in those months to be statistically meaningful.

4) Change over Time?



Graphing the differences by date does not show anything obvious. There do not seem to be any abrupt changes or repeated patterns that might suggest equipment calibration slippage (and recalibration).

5) Volunteer Variability?



Are some volunteer values closer than others? **Teams 4 and 11 have noticeably better confidence intervals** and fewer outliers. Teams are numbered here to preserve anonymity.

Note: "Small" and "N/A" are aggregates for individuals with few observations or were unidentified.

6) Sampling Technique Change?

The YSI 95 meter was replaced 7/1/05 with model 85. **The new meter required stirring for accurate DO.** The new volunteer sampling procedure was not immediately implemented by all

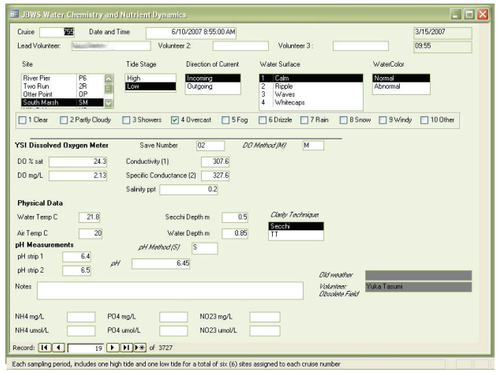
	Mean	Std Dev	95% Confidence Interval
Old meter	.448	1.354	2.65
New meter	-.28	.979	1.92
New meter, new method	**-.21**	**.922**	**1.80**
All data combined	.01	1.189	2.33

- **Volunteer data with the correct procedure more closely matches the automated data** (narrower confidence interval).
- There is a substantial difference between the variation in differences from the old meter and new meter (using the correct procedure).

Data Quality Protection Features

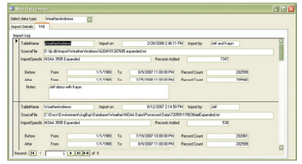
Data Entry and Editing

- Range check numeric values (physical limits, historical min, max, mean)
  Warns user of potential anomalies while correction may be easy
- Validate related data (e.g., DO mg/l versus % saturation)
  Can detect errors in values that individually are in range.
- Data screen corresponds to data sheet, and meter display sequence
  Quicker and less error prone data entry.
- Compute values (nutrient unit conversion, average pH)
  Eliminates risk of manual calculation error
- Have notes area for other info or data entry notes
  Stores information written on data sheet margins, etc.
- Use list selection when values are limited
  Increases consistency, prevents typographical errors
- Avoid manual data entry error by direct download instead of typing
  Direct data transfer from instrument, data logger, smart phone, etc. eliminates transcription errors… OR
- Both directly download and enter from datasheets.
  Provides secondary source of data if equipment or operator error results in loss of one version.



Data entry/review screen for volunteer water quality data.

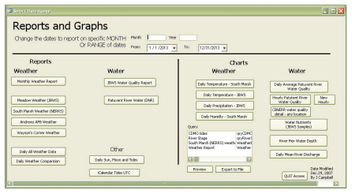
Data Integrity



Log showing results of import of data from external source.

- Warn user when changing existing record
  Reduces risk of saving accidental changes. MS Access automatically saves changes when the person goes to the next record. Efficient for entering new data, but
- Database has fields for future use (e.g. pH method is now always "strip")
  When method changes in the future, the existing data already have the correct method recorded
- Obsolete data fields are still visible but clearly marked
  Old values are available in case of discrepancy
- Automatically log data import metadata (file information, record count, date/time, person performing import, etc.)
  Provides traceability of data back to original electronic source to resolve discrepancies detected later.

Data Warehouse



A **repository for abiotic data** easily integrated with project-specific data including benthic macro-invertebrates, fish, aquatic turtle movement, and submerged and emergent vegetation.

Contents

- **Weather**
  4 stations-
  National Weather Service, NOAA NERRS/Maryland Dept. Nat'l Resources, Weather Underground, on-site station
- **River depth**
  (NERRS) and flow rate (USGS), predicted tides (NOAA)
- **Water Quality**
  Volunteer data , automated river data (NERRS/DNR), DNR grab samples -nutrients

Benefits

- Computes linear interpolation of data values from automated values before and after volunteer data times to compare with volunteer values.
- Allows easy sorting and grouping (month, volunteer, etc)
- Simple repetition of analysis with "what-if" conditions
- Export data for further analysis (e.g., Excel, R Statistics) or for GIS as KML or to become shape files.

Next Generation Analysis

Greater integration of meteorological, tidal, and upstream river stage data to quantify relationships and ecological impacts.

Examples

- Model of water depth at site based upon USGS and NWS data could be used to extrapolate site water levels prior to installation of automated monitoring. Estuarine inundation periods are an important ecological factor.
- Develop educational case studies of storm precipitation on river levels, turbidity, temperature to illustrate the significance of runoff issues to the Chesapeake Bay .

Planned Additions to the Data Warehouse

- Instrument calibration and maintenance log for interpretation of data uncertainty.
- "Watershed events" that can impact water quality such as sewage overflow at the nearby processing plant, excess road salt, or adoption of agricultural and storm-water best management practices upstream.
- Inclusion of existing history of benthic macroinvertebrate data.
- Initiate stream geomorphological monitoring to better interpret habitat change as a factor in variations in macroinvertebrate and fish survey results

Additional Detail on Data Management Recommendations

WWW.ENVIRONMENTALINFORMATICS.COM

- Information Technology > Improving Usability for Data Entry
- Environmental Data > QA/QC Field Data Hints
- Environmental Data > QA/QC Water Quality *(this poster)*

Acknowledgments

Lindsay Hollister and Robert Smith at Jug Bay Wetlands Sanctuary contributed historical perspectives and insights for interpretation of the initial analysis. We thank the multitude of volunteers who have completed their appointed rounds despite snow, rain, heat, and gloom of fog.
Continuous monitoring data from the Maryland Department of Natural Resources and NOAA National Estuarine Research Reserve System.